What is the difference between rollup and scan?

Ans: By using rollup we cant generate cumulative summary records for that we will be using scan.

What is the difference between partitioning with key and round robin?

Ans: PARTITION BY KEY:
In this, we have to specify the key based on which the partition will occur. Since it is key based it results in very well balanced data. It is useful for key dependent parallelism.
PARTITION BY ROUND ROBIN:In this, the records are partitioned in sequential way, distributing data evenly in blocksize chunks across the output partition. It is not key based and results in well balanced data especially with blocksize of 1. It is useful for record independent parallelism.

How do you truncate a table

ans: There are many ways to do it.
1. Probably the easiest way is to use Truncate Table
2. Run Sql or update table can be used to do the same thing
3. Run Program

What is the difference between a DB config and a CFG file?

Ans; A .dbc file has the information required for Ab Initio to connect to the database to extract or load tables or views. While .CFG file is the table configuration file created by db_config while using components like Load DB Table

Types of parallelism in detail.

ans:There are 3 types of parallelism in ab-initio.
1) Data Parallelism: Data is processed at the different servers at the same time.
2) Pipeline parallelism: In this the records are processed in pipeline, i.e. the components do not have to wait for all the records to be processed. The records that got processed are passed to next component in pipeline.
3) Component Parallelism: In this two or more components process the records in parallel.
Component parallelism:- A graph with multiple processes running simultaneously on
separate data uses component parallelism.
Data parallelism :- A graph that deals with data divided into segments and operates on
each segment simultaneously uses data parallelism. Nearly all commercial data processing tasks can use data parallelism. To support this form of parallelism, Ab Initio
provides Partition components to segment data, and Departition components to merge
segmented data back together .
Pipeline parallelism :- A graph with multiple components running simultaneously on the same data uses pipeline parallelism. Each component in the pipeline continuously

reads from upstream components, processes data, and writes to downstream components. Since a downstream component can process records previously written by an upstream component, both components can operate in parallel. NOTE: To limit the number of components running simultaneously, set phases in the graph.

What is the function you would use to transfer a string into a decimal?

Ans: For converting a string to a decimal we need to typecast it using the following syntax,

out.decimal_field :: (decimal(size_of_decimal)) string_field;

The above statement converts the string to decimal and populates it to the decimal field in output.

. How to execute the graph from start to end stages? Tell me and how to run graph in non-Abinitio system?

Ans: There are so many ways to do this, i am giving one example due to time constraint you can run components according to phasea how you defined.

by creating ksh, sh scripts also you can run.

. What is data mapping and data modelling?

Ans; Data mapping deals with the transformation of the extracted data at FIELD level i.e. the transformation of the source field to target field is specified by the mapping defined on the target field. The data mapping is specified during the cleansing of the data to be loaded.

For Example:

source;

string(35) name = "Siva Krishna ";

target;

string("01") nm=NULL("");/*(maximum length is string(35))*/

Then we can have a mapping like:

Straight move.Trim the leading or trailing spaces.

The above mapping specifies the transformation of the field nm

What is the difference between sandbox and EME, can we perform checkin

and checkout through sandbox/ Can anybody explain checkin and checkout?

Ans; Sandboxes are work areas used to develop, test or run code associated with a given project. Only one version of the code can be held within the sandbox at any time.

The EME Datastore contains all versions of the code that have been checked into it. A particular sandbox is associated with only one Project where as a Project can be checked out to a number of sandboxes

explain the environment varaibles with example.?

ans; Environemental variables server as global variables in unix envrionment. They are used for passing on values from a shell/ process to another. They are inherited by Abinitio as sandbox variables/ graph parameters like

AI_SORT_MAX_CORE

AI_HOME
AI_SERIAL
AI_MFS etc.
To know what all variables exist, in your unix shell, find out the naming convention and type a command like "env | grep AI". This will provide you a list of all the variables set in the shell. You can refer to the graph parameters/ components to see how these variables are used inside Abinitio.

What r the Graph parameter?
ans: There are 2 types of graph parameters in AbInitio
1. local parameter
2. Formal parameters.(those parameters working at runtime)

. How to Improve Performance of graphs in Ab initio?Give some examples or tips.?
Ans: There are somany ways to improve the performance of the graphs in Abinitio.
I have few points from my side.
1.Use MFS system using Partion by Round by robin.
2.If needed use lookup local than lookup when there is a large data.
3.Takeout unnecessary components like filter by exp instead provide them in reformat/Join/Rollup.
4.Use gather instead of concatenate.
5.Tune Max_core for Optional performance.
6.Try to avoid more phases.

What are the most commonly used components in a Abinition graph. example of a trasformation of data, say customer data in a credit card company into meaningful output based on business rules?
Ans: The most commonly used components in to any Ab Initio project are
input file/output file
input table/output table
lookup file
reformat,gather,join,runsql,join with db,compress components,sort,trash,partition by expression,partition by key ,concatinate

Difference between conventional loading and direct loading ? when it is used in real time .?
ans: Conventional Load:
Before loading the data, all the Table constraints will be checked against the data.
Direct load:(Faster Loading)
All the Constraints will be disabled. Data will be loaded directly.Later the data will be checked against the table constraints and the bad data won't be indexed.
Api conventional loading
utility direct loading.

How to find the number of arguments defined in graph?
Ans: $# - No of positional parameters
$? - the exit status of the last executed command.

. What is the difference between .dbc and .cfg file?
Ans: .cfg file is forÂ the remote connection and .dbc is for connecting the database.
.cfg contains :
1. The name of the remote machine
2. The username/pwd to be used while connecting to the db.
3. The location of the operating system on the remote machine.
4. The connection method.
and .dbc file contains the information:
1. The database name2. Database version
3. Userid/pwd
4. Database character set and some more...

. How to do we run sequences of jobs ,,like output of A JOB is Input to B .How
do we co-ordinate the jobs?
Ans: By writing the wrapper scripts we can control the sequence of execution of more
than one job.

How would you do performance tuning for already built graph ? Can you let me know
some examples?
Ans: example :- suppose sort is used in fornt of merge component its no use of using
sort ! bcz we hv sort component built in merge.
2) we use lookup instead of JOIN,Merge Componenet.
3) suppose we wnt to join the data comming from 2 files and we dnt wnt dupliates we
will use union funtion instead of adding addtional component for duplicate remover.

. What is semi-join
ans: In abinitio,there are 3 types of join...
1.inner join. 2.outer join and 3.semi join.
for inner join 'record_requiredn' parameter is true for all in ports.
for outer join it is false for all the in ports.
if u want the semi join u put 'record_requiredn' as true for the required component
and
false for other components..

How to get DML using Utilities in UNIX?
Ans: If your source is a cobol copybook, then we have a command in unix which
generates the required in Ab Initio. here it is:
cobol-to-dml.

what is local and formal parameter?

Ans: Two are graph level parameters but in local you need to initialize the value at the time of declaration where as globle no need to initialize the data it will promt at the time of running the graph for that parameter.

. what is BRODCASTING and REPLICATE ?
ans: Broadcast - Takes data from multiple inputs, combines it and sends it to all the output ports.
Eg - You have 2 incoming flows (This can be data parallelism or component parallelism) on Broadcast component, one with 10 records & other with 20 records. Then on all the outgoing flows (it can be any number of flows) will have 10 + 20 = 30 records
Replicate - It replicates the data for a particular partition and send it out to multiple out ports of the component, but maintains the partition integrity.
Eg - Your incoming flow to replicate has a data parallelism level of 2. with one partition
having 10 recs & other one having 20 recs. Now suppose you have 3 output flos from replicate. Then each flow will have 2 data partitions with 10 & 20 records respectively.

What is m_dump
m_dump command prints the data in a formatted way.
m_dump <dml> <file.dat>

an exaple of realtime start script in the graph?
Ans: Here is a simple example to use a start script in a graph:
In start script lets give as:
export $DT=`date '+%m%d%y'`
Now this variable DT will have today's date before the graph is run.
Now somewhere in the graph transform we can use this variable as;
out.process_dt::$DT;
which provides the value from the shell.

How to run the graph without GDE?
Ans: In RUN ==> Deploy >> As script , it create a .bat file at ur host directory ,and then run .bat file from Command prompt

How Does MAXCORE works?
Ans: Maxcore is a value (it will be in Kb).Whne ever a component is executed it will take that much memeory we specified for execution

.What is $mpjret? Where it is used in ab-initio?
ans: You can use $mpjret in endscript like
if 0 -eq($mpjret)then
echo "success"
else
mailx -s "[graphname] failed" mailid

How do you convert 4-way MFS to 8-way mfs?

Ans: To convert 4 way to 8 way partition we need to change the layout in the partioning component. There will be seperate parameters for each and every type of partioning eg. AI_MFS_HOME, AI_MFS_MEDIUM_HOME, AI_MFS_WIDE_HOME etc. The appropriate parameter need to be selected in the component layout for the type of
partioning..

What is AB_LOCAL expression where do you use it in ab-initio?

ans: ablocal_expr is a parameter of itable component of Ab Initio.ABLOCAL() is replaced by the contents of ablocal_expr.Which we can make use in parallel unloads.There are two forms of AB_LOCAL() construct, one with no arguments and one with single argument as a table name(driving table).
The use of AB_LOCAL() construct is in Some complex SQL statements contain grammar that is not recognized by the Ab Initio parser when unloading in parallel. You can use the ABLOCAL() construct in this case to prevent the Input Table component from parsing the SQL (it will get passed through to the database). It also specifies which table to use for the parallel clause.

What is mean by Co > Operating system and why it is special for Abinitio ?

ans: It converts the AbInitio specific code into the format, which the UNIX/Windows can understand and feeds it to the native operating system, which carries out the task.

How will you test a dbc file from command prompt ?

ans: try "m_db test myfile.dbc"

. Which one is faster for processing fixed length dmls or delimited dmls and why ?

ans: Fixed length DML's are faster because it will directly read the data of that length without any comparisons but in delimited one,s every character is to be compared and
hence delays

.What are the continuous components in Abinitio?

ans: Contineous components used to create graphs,that produce useful output file while running continously
Ex:- Contineous rollup,Contineous update,batch subscribe

How to retrieve data from database to source in that case whice component is used for this?

ans; To unload (retrive) Data from the database DB2, Informix, or Oracle we have components like Input Table and Unload DB Table by using these two components we can unload data from the database.

. What is the relation between EME , GDE and Co-operating system ?
ans: EME is said as enterprise metdata env, GDE as graphical devlopment env and Cooperating sytem can be said as asbinitio server relation b/w this CO-OP, EME AND GDE
is as fallows
    Co operating system is the Abinitio Server. this co-op is installed on perticular O.S platform that is called NATIVE O.S .comming to the EME, , its hold the metadata,trnsformations,db config files source and targets information's. comming to GDE its is end user envirinment where we can devlop the graphs(mapping just like in informatica)
designer uses the GDE and designs the graphs and save to the EME or Sand box it is at user side.where EME is ast server side.

. What are kinds of layouts does ab initio supports
ans: Basically there are serial and parallel layouts supported by AbInitio. A graph can have both at the same time. The parallel one depends on the degree of data parallelism. If the multi-file system is 4-way parallel then a component in a graph can run 4 way parallel if the layout is defined such as it's same as the degree of parallelism.

Do you know what a local lookup is?
ans: Lookup File consists of data records which can be held in main memory. This makes the transform function to retrieve the records much faster than retirving from disk. It allows the transform component to process the data records of multiple files fastly.

. How many components in your most complicated graph?
ans: This is a tricky question, number of component in a graph has nothing to do withthe level of knowledge a person has. On the contrary, a proper standardized and modular parametric approach will reduce the number of components to a very few. In a well thought modular and parametric design, mostly the graphs will have 3/4 components, which will be doing a particular task and will then call another sets of graphs to do the next and so on. This way total numbers of distinct graphs will drastically come down, support and maintenance will be much more simplified.
The bottomline is, there are lot more other things to plan rather than to add components.

. How to handle if DML changes dynamically in abinitio
ans: If the DML changes dynamically then both dml and xfr has to be passed as graph level parameter during the runtime.

. Have you worked with packages?
Ans: Packages are nothing but the reusable blocks of objects like transforms, user defined functions, dmls etc. These packages are to be included in the transform where you use them. For example, consider a user defined function like
/*string_trim.xfr*/

```
out::trim(input_string)=
begin
let string(35) trimmed_string = string_lrtrim(input_string);
out::trimmed_string;
end
```
Now, the above xfr can be included in the transform where you call the above function
as
include ''~/xfr/string_trim.xfr'';
But this should be included ABOVE your transform function.
For more details see the help file in "packages".

. What are primary keys and foreign keys?

Ans: In RDBMS the relationship between the two tables is represented as Primary key
and foreign key relationship.Wheras the primary key table is the parent table and
foreignkey table is the child table.The criteria for both the tables is there should be a
matching column.

What are Cartesian joins?

Ans: Cartesian join will get you a Cartesian product. A Cartesian join is when you join
every row of one table to every row of another table. You can also get one by joining
every row of a table to every row of itself.

Explain the difference between the "truncate" and "delete" commands?

ans: Truncate :- It is a DDL command, used to delete tables or clusters. Since it is a
DDL command hence it is auto commit and Rollback can't be performed. It is faster
than delete.

. How can i run the 2 GUI merge files?

Ans:Do you mean by merging Gui map files in WR.If so, by merging GUI map files in
GUI map editor it wont create corresponding test script.without testscript you cant run
a file.So it is impossible to run a file by merging 2 GUI map files.